From *SIAM News*, Volume 45, Number 9, November 2012

Uncertainty Quantification 2012

Uncertainty Quantification for Environmental Models

By Mary C. Hill, Dmitri Kavetski, Martyn Clark, Ming Ye, and Dan Lu

Environmental models are used to evaluate the fate of fertilizers in agricultural settings (including soil denitrification), the degradation of hydrocarbons at spill sites, and water supply for people and ecosystems in small to large basins and cities—to mention but a few applications of these models. They also play a role in understanding and diagnosing potential environmental impacts of global climate change. The models are typically mildly to extremely nonlinear. The persistent demand for enhanced dynamics and resolution to improve model realism [17] means that lengthy individual model execution times will remain common, notwithstanding continued enhancements in computer power. In addition, high-dimensional parameter spaces are often defined, which increases the number of model runs required to quantify uncertainty [2]. Some environmental modeling projects have access to extensive funding and computational resources; many do not.

The many recent studies of uncertainty quantification in environmental model predictions have focused on uncertainties related to data error and sparsity of data, expert judgment expressed mathematically through prior information, poorly known parameter values, and model structure (see, for example, [1,7,9,10,13,18]). Approaches for quantifying uncertainty include frequentist (potentially with prior information [7,9]), Bayesian [13,18,19], and likelihood-based. A few of the numerous methods, including some sensitivity and inverse methods with consequences for understanding and quantifying uncertainty, are as follows: Bayesian hierarchical modeling and Bayesian model averaging; single-objective optimization with error-based weighting [7] and multi-objective optimization [3]; methods based on local derivatives [2,7,10]; screening methods like OAT (one at a time) and the method of Morris [14]; FAST (Fourier amplitude sensitivity testing) [14]; the Sobol' method [14]; randomized maximum likelihood [10]; Markov chain Monte Carlo (MCMC) [10]. There are also bootstrapping and cross-validation approaches.Sometimes analyses are conducted using surrogate models [12].

The availability of so many options can be confusing. Categorizing methods based on fundamental questions assists in communicating the essential results of uncertainty analyses to stakeholders. Such questions can focus on model adequacy (e.g., How well does the model reproduce observed system characteristics and dynamics?) and sensitivity analysis (e.g., What parameters can be estimated with available data? What observations are important to parameters and predictions? What parameters are important to predictions?), as well as on the uncertainty quantification (e.g., How accurate and precise are the predictions?).

The methods can also be classified by the number of model runs required: few (10s to 1000s) or many (10,000s to 1,000,000s). Of the methods listed above, the most computationally frugal are generally those based on local derivatives; MCMC methods tend to be among the most computationally demanding. Surrogate models (emulators)do not necessarily produce computational frugality because many runs of the full model are generally needed to create a meaningful surrogate model. With this categorization, we can, in general, address all the fundamental questions mentioned above using either computationally frugal or demanding methods. Model development and analysis can thus be conducted consistently using either computationally frugal or demanding methods; alternatively, different fundamental questions can be addressed using methods that require different levels of effort.

Based on this perspective, we pose the question: Can computationally frugal methods be useful companions to computationally demanding methods? The reliability of computationally frugal methods generally depends on the model being reasonably linear, which usually means smooth nonlinearities and the assumption of Gaussian errors; both tend to be more valid with more linear models. The reliability of computationally demanding methods depends on wise choice of parameter-value ranges and on a sufficient number and proper distribution of parameter samples. Many theoretical and empirical comparisons suggest that frugal computational methods often produce results similar to those for computationally demanding methods [9], indicating that in many circumstances nonlinearities may not be as problematic as sometimes feared. Figure 1 compares uncertainty intervals calculated with computationally frugal linear and nonlinear confidence intervals, and with demanding MCMC credible intervals. The problem is synthetic, which means that the true value of the prediction is known. For this problem, it appears that difficulties caused by model inadequacy are more serious than the approximations made in order to use computationally frugal instead of computationally demanding UQ methods. This suggests that at times, a wise approach may be to use mainly computationally frugal UQ methods and to focus resources on exploring alternative models.

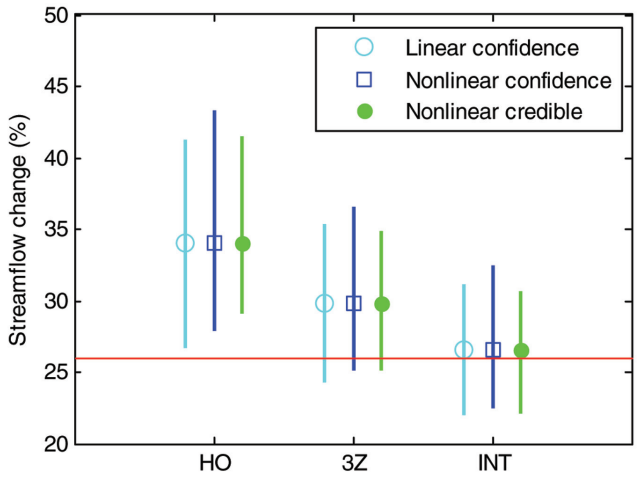

Figure 1. *Linear and nonlinear confidence intervals and nonlinear credible intervals (using INT as an example; 106, 1594, and 420,000 model runs, respectively) on predicted change in streamflow caused by pumpage for three alternative models. The horizontal line defines the true value of the prediction, which is known for this synthetic problem. The nonlinear credible intervals are calculated by MCMC with a DREAM algorithm. (After [9].)*

Recent investigations of model nonlinearity have suggested that models can be more nonlinear than the systems they attempt to replicate. In [8] and references cited therein, thresholds are discussed as a source of unrealistic nonlinearity. Commonly, below a thresh-

old value for simulated results, a variable is held constant, while above the threshold linear variations occur. Thresholds may be consistent with small-scale results, yet averaging mechanisms in complex environments may justify a smoother function. Using a smoother curve profoundly affects the reliability of frugal methods based on local derivatives. Similarly, erratic performance of time-stepping routines can produce dramatic, and unrealistic, model nonlinearities that deteriorate the reliability of computationally frugal methods of uncertainty evaluation. Making models more robust by eliminating false nonlinearities and numerical artifacts makes it easier to understand real system nonlinearities. The greater understanding is derived in part because the frugality of the computations allows analyses at multiple sets of parameter values. For example, graphs produced for different sets of parameter values (see Figure 2) showed that though the same parameters remained important, the dominant observation types changed. Such insight can be important to data-monitoring decisions and can be obscured by false linearities.

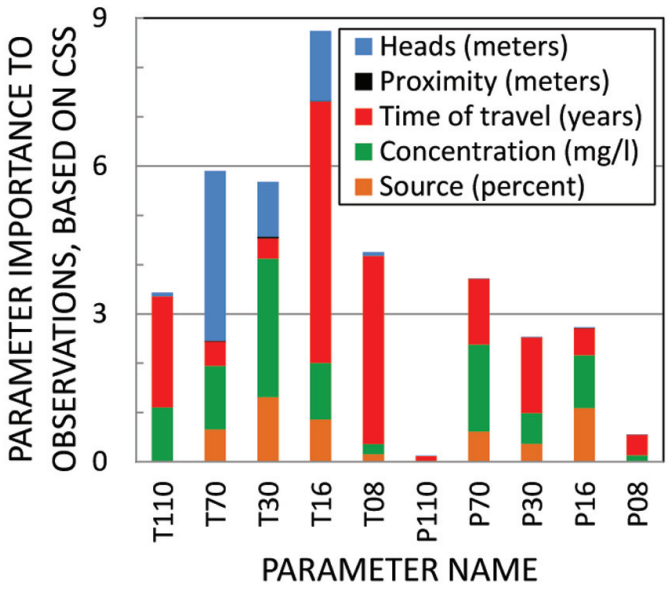

Figure 2. *Parameter importance to observed quantities or, conversely, the information content from different observation types for the listed parameters. In this groundwater problem, T parameters are transmissivities and P parameters are porosities. The key lists observation types considered: hydraulic heads (a measure of potential energy); proximity of transported particles to a defined location at a defined time; the time it takes particles to travel between defined areas; the concentration of perchloroethylene and chlorofluorocarbon at defined locations and times; and the source of water reaching pumped wells. Parameter correlation coefficients (PCCs) showed little interdependence between parameters. The results shown require 21 parallelizable model runs. Results are similar for many sets of parameter values considered during the course of model calibration, suggesting that nonlinearity, though considerable for this problem, is not debilitating. (Modified from [4].)*

The results shown in Figures 2 and 3 address two of the fundamental questions posed earlier. Figure 2 offers insight into the questions: What parameters can be estimated with available data? What observations are important to parameters? These questions can be answered with computationally demanding methods, such as MCMC and FAST, as well as computationally frugal methods, one of which is shown here [2,7]. The key to reliable results for the statistic shown is careful scaling, as discussed in [7].

The composite scaled sensitivity (CSS) and parameter correlation coefficient (PCC) are calculated as follows [7]:

$$\mathrm{CSS}_k = \{\Sigma_{i=1,n}\,[\Sigma_{j=1,n}\,(\partial y'_k/\partial b_j) b_j (\omega^{1/2})_{ki}]^2\}^{1/2}, \qquad k = 1,np; \tag{1}$$

$$\mathrm{PCC}_{k,j} = \boldsymbol{V}_{k,j}\,(\boldsymbol{b})/[V_{j,j}\,(\boldsymbol{b})^{1/2} V_{k,k}\,(\boldsymbol{b})^{1/2}]\;\; \boldsymbol{V}_{kj}\,(\boldsymbol{b}) = [s^2(\boldsymbol{X}^T\omega\boldsymbol{X})^{-1}]_{k,j}, \quad k = 1,np; \quad j = 1,np. \tag{2}$$

For the results presented in this work, n is the number of observations; $(\partial y'_k/\partial b_j)$ the sensitivity of the kth simulated value y'_k to the jth parameter b_j; and np the number of parameters. Sensitivities were calculated by MODFLOW-2000 [5] with the sensitivity-equation method in $np + 1$ model runs, or by perturbation with central differences in $(2 \times np) + 1$ model runs using UCODE_2005 [11]. $\boldsymbol{V}_{i,j}(\boldsymbol{b})$ is the entry in the parameter variance–covariance matrix for parameters i and j; this is a variance for $i = j$, a covariance for $i \neq j$. $\boldsymbol{X}$ is a matrix of sensitivities with entries equal to $(\partial y'_k/b_j)$, ω is the weight matrix, and s^2 is the unbiased regression variance. PCC for extremely correlated parameters can be calculated through creative use of round-off error [6,7].

Figure 3 addresses the question: What observations are important to predictions? The importance of existing old observations and potential new observations is considered. For both, the importance of different observations depends on choices made in model construction, and results shown in Figure 3 reveal consequences of such decisions. The computationally frugal observation–prediction (OPR) statistic used is defined as how much a calculated confidence interval would increase if existing observations were removed and how much it would decrease if new observations were added [16]. The equations are:

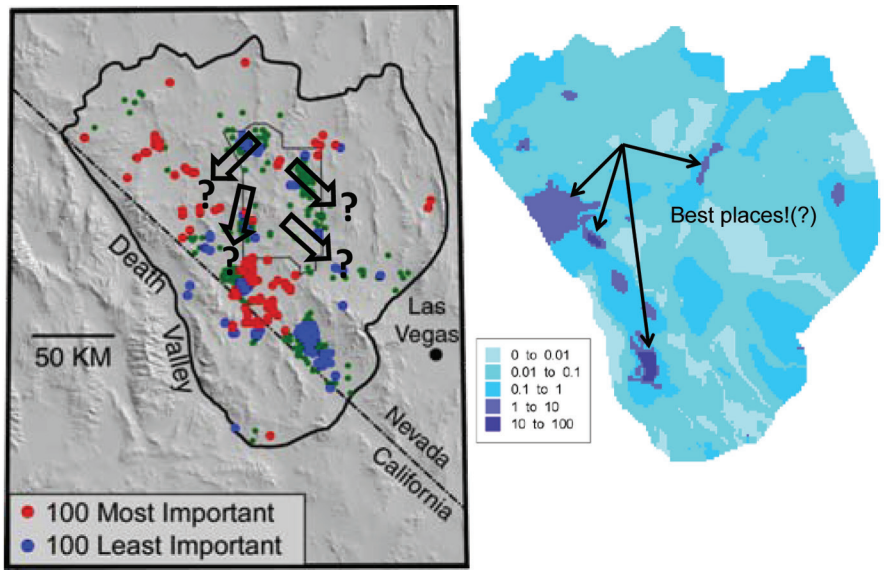

Figure 3. *Importance of observations to predictions of transport within the Nevada National Security Site (NNSS). Left, the NNSS is outlined in gray and the model boundary in black, and the transport locations are represented schematically. The observation–prediction (OPR) statistic is used to measure observation importance [16]. The existing old 501 hydraulic head observations are ranked. Right, evaluation of one potential new head measurement anywhere in model layer 1. The most important observations in the southwestern part of the model occur largely because the rocks there are defined in the model as hydraulically similar to the rocks in the NNSS, but their occurrence here under steep head gradients facilitates estimation of the parameter value. Each of these results required 49 parallelizable model runs. (Modified from [7] and [15].)*

$$\mathrm{OPR}_i = 100 \times (s_{z(i)} - s_z)/s_z \tag{3a}$$

$$s_{z(i)} = [(\partial z/\partial \boldsymbol{b})^T [s^2 (\boldsymbol{X}_{(i)}{}^T \boldsymbol{\omega}_{(i)} \boldsymbol{X}_{(i)})^{-1}] \, (\partial z/\partial \boldsymbol{b})]^{1/2} \tag{3b}$$

$$s_z = [(\partial z/\partial \boldsymbol{b})^T [s^2 (\boldsymbol{X}^T \boldsymbol{\omega} \boldsymbol{X})^{-1}] \, (\partial z/\partial \boldsymbol{b})]^{1/2}, \tag{3c}$$

where i identifies the observation, and $\boldsymbol{X}_{(i)}$ and $\boldsymbol{\omega}_{(i)}$ indicate that the sensitivity matrix $\boldsymbol{X}$ and weight matrix $\boldsymbol{\omega}$ have been modified by the addition of rows and columns related to new observation i. The importance of existing observations is evaluated by the removal of rows and columns of $\boldsymbol{X}$ and $\boldsymbol{\omega}$, as indicated by $(-i)$; in practice, entries in the weight matrix are set to zero. The weight matrix is determined through an analysis of errors, as required to obtain minimum variance parameter estimates ([7], Appendix C). The use of a standard deviation in equation (3) instead of a confidence-interval width is consistent with an assumed Gaussian distribution.

The brief analysis and references presented here suggest that increasingly, as models become more robust, a full uncertainty toolbox that includes computationally frugal methods, such as methods based on local derivatives, along with computationally demanding methods, such as MCMC, and presentation of results in the context of fundamental questions in ways that facilitate comparisons between different models and hypotheses, will best serve the needs of environmental modeling.

References

[1] R.C. Aster, B. Borshers, and C.H. Thurber, *Parameter Estimation and Inverse Problems*, Academic Press, Amsterdam, 2012.

[2] J. Doherty, *PEST*, Watermark Computing, Brisbane, Australia, 2012.

[3] A. Efstratiadis and D. Koutsoyiannis, *One decade of multi-objective calibration approaches in hydrological modelling: A review*, Hydrol. Sci. J., 55:1 (2010), 58–78.

[4] R.T. Hanson, L.K. Kauffman, M.C. Hill, J.E. Dickinson, and S.W. Mehl, *Advective Transport Observations with MODPATH-OBS—Documentation of the MODPATH Observation Process Using Four Types of Observations and Predictions,* in *U.S. Geological Survey Techniques and Method*s, 6–A42, 2012.

[5] M.C. Hill, E.R. Banta, A.W. Harbaugh, and E.R. Anderman, *MODFLOW-2000, the U.S. Geological Survey modular ground-water model: User's guide to the observation, sensitivity, and parameter-estimation process and three post-processing programs*, U.S. Geological Survey Open-File Report 00–184, 2000, http://water.usgs.gov/nrp/gwsoftware/modflow2000/modflow2000.html.

[6] M.C. Hill and O. Østerby, *Determining extreme parameter correlation in ground-water models,* Ground Water, 41:4 (2003), 420–430.

[7] M.C. Hill and C.R. Tiedeman, *Effective Calibration of Ground Water Models, with Analysis of Data, Sensitivities, Predictions, and Uncertainty*, John Wiley & Sons, Hoboken, NJ, 2007.

[8] D. Kavetski and M.P. Clark, *Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction*, Water Resour. Res., 46:W10511 (2010), doi:10.1029/2009WR008896.

[9] D. Lu, M. Ye, and M.C. Hill, *Analysis of regression confidence intervals and Bayesian credible intervals for uncertainty quantification*, Water Resour. Res., 48:W0951 (2012), doi:10.1029/2011WR011289.

[10] D.S. Oliver, A.C. Reynolds, and N. Liu, *Inverse Theory for Petroleum Reservoir Characterization and History Matching*, Cambridge University Press, UK, and New York, 2008.

[11] E.P. Poeter, M.C. Hill, E.R. Banta, S. Mehl, and S. Christensen, *UCODE_2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation*, in *U.S. Geological Survey Techniques and Methods*, 6–A11, 2005, http://typhoon.mines.edu/freeware/ucode/.

[12] S. Razavi, B.A. Tolson, and D.H. Burns, *Review of surrogate modeling in water resources*, Water Resour. Res., 48:W07401 (2012), doi:10.1029/2011WR011527.

[13] B. Renard, D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S.W. Franks, *Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation*, Water Resour. Res., 47:W11516 (2011), doi:10.1029/2011WR010643.

[14] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, et al., *Global Sensitivity Analysis: The Primer*, John Wiley & Sons, Hoboken, NJ, 2008.

[15] C.R. Tiedeman, D.M. Ely, M.C. Hill, and G.M. O'Brien, *A method for evaluating the importance of system state observations to model predictions, with application to the Death Valley regional groundwater flow system*, Water Resour. Res., 40:W12411 (2004), doi:10.1029/2004WR003313.

[16] M.J. Tonkin, C.R. Tiedeman, D.M. Ely, and M.C. Hill, *OPR-PPR, a computer program for assessing data importance to model predictions using linear statistics*, in *U.S. Geological Survey Techniques and Methods,* 6–E2, 2007, http://water.usgs.gov/software/OPR-PPR.html.

[17] J.A. Vrugt, C.J.F. ter Braak, M.P. Clark, J.M. Hyman, and B.A. Robinson, *Treatment of input uncertainy in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation*, Water Resour. Res., 44 W00B09 (2008), doi:10.1029/2007WR006720.

[18] E.F. Wood, et al., *Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water*, Water Resour. Res., 47:W05301 (2011), doi:10.1029/2010WR010090.

[19] M. Ye, P.D. Meyer, and S.P. Neuman, *On model selection criteria in multimodel analysis*, Water Resour. Res., 44:W03428 (2008), doi:10.1029/2008WR006803.

Acknowledgments

The references were selected from areas of environmental modeling to form an introduction for interested readers. A comprehensive list would be inordinately long for this publication; additional information can be found in works cited in some of the references listed.

The authors thank Jeremy White of the U.S. Geological Survey and Luis Tenorio of the Colorado School of Mines for their reviews of this article.

Mary Hill was funded by the U.S. Geological Survey programs NAWQA (National Water Quality Assessment), GWRP (Groundwater Resources Program), and NRP (National Research Program). Ming Ye and Dan Lu were funded by NSF–EAR grant 0911074 and DOE Early Career Award DE–SC0008272.

Mary C. Hill is a senior research hydrologist at the U.S. Geological Survey. Dmitri Kavetski is a professor of civil and environmental engineering at the University of Adelaide. Martyn Clark is a scientist at the National Center for Atmospheric Research. Ming Ye and Dan Lu are an associate professor and a postdoctoral fellow, respectively, in the Department of Scientific Computing at Florida State University.